US-PAT-NO: 5675706

DOCUMENT-IDENTIFIER: US 5675706 A

TITLE: Vocabulary independent discriminative utterance verification for non-keyword rejection in subword based speech recognition

DATE-ISSUED: October 7, 1997

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Lee; Chin-Hui	New Providence	NJ	N/A	N/A
Sukkar; Rafid Antoon	Aurora	IL	N/A	N/A

US-CL-CURRENT: 704/256, 704/254

ABSTRACT:

A verification system to determine unknown input speech contains a recognized keyword or consists of speech or other sounds that do not contain any of the keywords. The verification system is designed to operate on the subword level, so that the verification process is advantageously vocabulary independent. Such a vocabulary-independent verifier is achieved by a two-stage verification process comprising subword level verification followed by string level verification. The subword level verification stage verifies each subword segment in the input speech as determined by an Hidden Markov Model recognizer to determine if that segment consists of the sound corresponding to the subword that the HMM recognizer assigned to that segment. The string level verification stage combines the results of the subword level verification to make the rejection decision for the whole keyword. Advantageously, the training of this two-stage verifier is independent of the specific vocabulary set implying that when the vocabulary set is update or changed the verifier need not be retrained and can still be reliably verifying the new set of keywords.

14 Claims, 10 Drawing figures

Exemplary Claim Number: 1

Number of Drawing Sheets: 5

---------- KWIC ---------

Detailed Description Text - DETX (32):

According to this invention the concept of "anti-subword **class**" HMM models is used. The anti-subword **class** models are constructed by, first, clustering the subword units into J **classes,** where J<<K. A hierarchical clustering algorithm is used to cluster subwords based on minimizing the overall inter-subword cluster confusion rate. Subwords that are highly **confusable** with other subwords are likely to be clustered together. Given the subword model set, [s.sub.i], the subword confusion information is obtained from the subword confusion matrix of the training set of sample sentences. Setting J=6, the constituency of each **class** is given in Table 5, which shows, for example, that certain vowel sounds are clustered into one **class (Class** B) while nasals are included under **Class** E. For each subword **class,** an anti-subword **class** model is

trained using all speech segments corresponding to sounds that are not modeled by any of the subwords in that subword **class**. Based on the **classes** shown in Table 5, a total of 6 anti-subword **class** models are constructed. The anti-subword model, s.sub.j, is now considered to be the anti-subword **class** model corresponding to the **class** to which s.sub.j belongs.

US-PAT-NO: 6487532

DOCUMENT-IDENTIFIER: US 6487532 B1

TITLE: Apparatus and method for distinguishing similar-sounding utterances speech recognition

DATE-ISSUED: November 26, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Schoofs; Koen	Haasrode	N/A	N/A	BE
Gallopyn; Guido	Zottegem	N/A	N/A	BE

US-CL-CURRENT: 704/251, 704/246

ABSTRACT:

Speaker-specified hints are used to establish conditions for a speech recognition system to select a recognition result for a previously provided utterance from among various possible homophones. The hints may characterize the utterance by a linguistic property, such as an orthographic, morphological, or semantic property.

25 Claims, 6 Drawing figures

Exemplary Claim Number: 1

Number of Drawing Sheets: 6

---------- KWIC ---------

Detailed Description Text - DETX (4):

The raw spectral information obtained from the front end circuitry 20 is further preprocessed in the computer 23 to replace each sample or input frame with an **index** which corresponds to or identifies one of a predetermined set of standard or prototype spectral distributions or frames. In the particular embodiment being described, 1024 such standard frames are utilized. In the art, this substitution is conventionally referred to as vector quantization and the **indices** are commonly referred to as VQ **indices**. The preprocessing of the input data by the computer 23 also includes an estimating of the beginning and end of a word or continuous phrase in an unknown speech input segment, e.g. based on the energy level values. For this purpose, the input circuitry may incorporate a software adjustable control parameter, designated the "sensitivity" value, which sets a threshold distinguishing user speech from background noise.

Detailed Description Text - DETX (5):

Vocabulary models are represented by sequences of standard or prototype states, which are represented by **indices**. Rather than representing spectral distributions, the state **indices** identify or correspond to probability distribution functions. The state spectral **index** essentially serves as a **pointer** into a table which identifies, for each state **index**, the set of probabilities that each prototype frame or VQ **index** will be observed to correspond to that state **index**. The table is, in effect, a precalculated

mapping between all possible frame **indices** and all state **indices**. Thus, for comparing a single frame and single state, a distance measurement or a measure of match can be obtained by directly indexing into the tables using the respective **indices** and combining the values obtained with appropriate weighting. It is thus possible to build a table or array storing a distance metric representing the closeness of match of each standard or prototype input frame with each standard or prototype model state. The distance or likelihood values which fill the tables can be generated by statistical training methods. A preferred system for precalculating and storing a table of distance measurements is disclosed in U.S. Pat. No. 5,546,499. The disclosure of that application is incorporated herein by reference.

Detailed Description Text - DETX (7):
In isolated word recognition, the sequence of frames which constitute the unknown speech segment taken together with a sequence of states representing a vocabulary model in effect define a matrix and the time warping process involves finding a path across the matrix which produces the best score, e.g., least distance or cost. The distance or cost is typically arrived at by accumulating the cost or distance values associated with each pairing of frame **index** with state **index** as described previously with respect to the VQ (vector quantization) process. An isolated word speech recognition system will typically identify the best scoring model and may also identify a ranked list of possible alternates.

Detailed Description Text - DETX (33):
The alternative filtering may be achieved in various manners. In one approach, all alternatives which lack the characters described in X are removed. Some pre-processing may be done, for example, "double L" would be replaced by "LL", then all alternatives not containing "LL" are removed. Another filtering approach exploits the fact that many hints are related to verb endings ("sent with a t"). Accordingly, the system may check whether the last letter(s) of the verb correspond to X. In this manner, X can be restrained to commonly confusable verb endings (e.g., d, t for English; e, s, es, t, ent for French). In another filtering approach, identifiers in a dictionary may be utilized to show to which letter a hint applies, if present (an **index** to a start position in the word string would suffice). For example, to differentiate KANJI characters, the hint may be stored in the dictionary entry for a word, such as in a field indicating the number of strokes in the character.

Detailed Description Text - DETX (35):
A preferred embodiment also has language model and grammar implications. In speech recognition, a word or a command can only be recognized if it is part of a grammar of a language model. This also applies to the hints as used in a preferred embodiment. Different options are possible to add hints to a language model. For example, the hint phrase "spelled with" may be modeled in the same way as a "capitalize that" command. That is, the hint can occur at any point in the dictation, after any word. This can be modeled by giving the hint a unigram occurrence probability. The value of the probability should be in line with the probability assigned to other commands such as "capitalize that". Alternatively, "spelled with" may be constrained to occurring only after certain **classes of confusable** words; for example, only after verbs.

US-PAT-NO: 5202952

DOCUMENT-IDENTIFIER: US 5202952 A

TITLE: Large-vocabulary continuous speech prefiltering and processing system

DATE-ISSUED: April 13, 1993

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Gillick; Laurence S.	Newton	MA	N/A	N/A
Roth; Robert S.	Newtonville	MA	N/A	N/A

US-CL-CURRENT: 704/200

ABSTRACT:

A continuous speech prefiltering system for use in continuous speech recognition computer systems. The speech to be recognized is converted from utterances to frame data sets, which frame data sets are smoothed to generate a smooth frame model over a predetermined number of frames. A resident vocabulary is stored within the computer as clusters of word models which are acoustically similar over a succession of frame periods. A cluster score is generated by the system, which score includes the likelihood of the smooth frames evaluated using a probability model for the cluster against which the smooth frame model is being compared. Cluster sets having cluster scores below a predetermined acoustic threshold are removed from further consideration. The remaining cluster sets are unpacked for determination of a word score for each unpacked word. These word scores are used to identify those words which are above a second predetermined threshold to define a word list which is sent to a recognizer for a more lengthy word match. A controller enables the system to initialize times corresponding to the frame start time for each frame data set, defining a sliding window.

27 Claims, 3 Drawing figures

Exemplary Claim Number: 22

Number of Drawing Sheets: 2

---------- KWIC ---------

Brief Summary Text - BSTX (11):

Continuous speech computational requirements are even greater. In continuous speech, the type of which humans normally speak, words are run together, without pauses or other simple cues to indicate where one word ends and the next begins. When a mechanical speech recognition system attempts to recognize continuous speech, it initially has no way of identifying those portions of speech which correspond to individual words. Speakers of English apply a host of duration and coarticulation rules when combining **phonemes** into words and sentences, employing the same rules in recognizing spoken language. A speaker of English, given a phonemic spelling of an unfamiliar word from a dictionary, can pronounce the word recognizably or recognize the word when it is spoken. On the other hand, it is impossible to put together an "alphabet" of recorded **phonemes** which, when concatenated, will sound like natural English

words. It comes as a surprise to most speakers, for example, to discover that the vowels in "will" and "kick", which are identical according to dictionary pronunciations, are as different in their spectral characteristics as the vowels in "not" and "nut", or that the vowel in "size" has more than twice the duration of the same vowel in "seismograph".

Detailed Description Text - DETX (3):

FIG. 1 is a general flow diagram showing the flow of information or data of the present invention. As shown, Phase I involves the flow of data from the user, in the form of utterances UT, through a series of transformers into transform data TR. The transform data is concurrently sent to a recognizer R and a processing, or pre-filter system PF. While the recognizer R processes the transform data TR, it **queries** the pre-filter system PF for data. Phase II involves the flow of transform data TR to the pre-filter system PF. Phase III then involves data flow of pre-filter data to a recognizer R upon **query** by the recognizer R. User U receives recognizer data in the form of a monitor word display on a monitor M. Each phase will separately be discussed below. The system of the present invention is used during Phase II for converting transform data into pre-filter data which is then sent for more lengthy filtering at a recognizer (Phase III).

Other Reference Publication - OREF (3):

Bahl et al., IEEE, Feb., 1990, pp. 85-88, "Constructing Groups of Acoustically **Confusable** Words".

Terms used **token indexing**

Found **6** of **661,433**

Sort results by: relevance

Display results: expanded form

Save results to a Binder

Search Tips

☐ Open results in a new window

Try an Advanced Search

Try this search in The Digital Library

Results 1 - 6 of 6

Relevance scale

1 Frame memory: a storage architecture to support rapid design and implementation of efficient databases
Salvatore T. March, Dennis G. Severance, Michael Wilens
September 1981 **ACM Transactions on Database Systems (TODS)**, Volume 6 Issue 3
Full text available: pdf(1.43 MB) Additional Information: full citation, abstract, references, citings, index terms

Frame memory is a virtual view of secondary storage that can be implemented with reasonable overhead to support database record storage and accessing requirements. Frame memory is designed so that its operating characteristics can be easily manipulated by either designers or design algorithms, while performance effects of such changes can be accurately predicted. Automated design procedures exist to generate and evaluate alternative database designs built upon frame memory, and the existenc ...

Keywords: analytic modeling, database design system, virtual secondary storge

2 RFC2967: TISDAG - Technical Infrastructure for Swedish Directory Access Gateways
L. Daigle, R. Hedberg
October 2000 rfc, RFC Editor
Additional Information: full citation

The strength of the TISDAG (Technical Infrastructure for Swedish Directory Access Gateways) project's DAG proposal is that it defines the necessary technical infrastructure to provide a single-access- point service for information on Swedish Internet users. The resulting service will provide uniform access for all information -- the same level of access to information (7x24 service), and the same information made available, irrespective of the service provider responsible for m ...

3 A two-view approach to constructing user interfaces
Gideon Avrahami, Kenneth P. Brooks, Marc H. Brown
July 1989 **ACM SIGGRAPH Computer Graphics , Proceedings of the 16th annual conference on Computer graphics and interactive techniques**, Volume 23 Issue 3
Full text available: pdf(4.40 MB) Additional Information: full citation, references, citings, index terms

4 A formal protection model of security in centralized, parallel, and distributed systems
Glenn S. Benson, Ian F. Akyildiz, William F. Appelbe
August 1990 **ACM Transactions on Computer Systems (TOCS)**, Volume 8 Issue 3
Full text available: pdf(2.17 MB) Additional Information: full citation, abstract, references, citings, index

terms, review

One way to show that a system is not secure is to demonstrate that a malicious or mistake-prone user or program can break security by causing the system to reach a nonsecure state. A fundamental aspect of a security model is a proof that validates that every state reachable from a secure initial state is secure. A sequential security model assumes that every command that acts as a state transition executes sequentially, while a concurrent security model assumes that multiple commands execut ...

Keywords: access control, concurrency control, distributed system security, operating system security, protection model

5 Musical information retrieval using melodic surface
Massimo Melucci, Nicola Orio
August 1999 **Proceedings of the fourth ACM conference on Digital libraries**
Full text available: pdf(674.04 KB) Additional Information: full citation, references, citings, index terms

Keywords: automatic indexing, automatic melodic segmentation, computer music, information retrieval, musical digital libraries

6 A storage and access manager for ill-structured data
Jeffrey Kottemann, Michael Gordon, Jack Stott
August 1991 **Communications of the ACM**, Volume 34 Issue 8
Full text available: pdf(2.04 MB) Additional Information: full citation, abstract, references, index terms, review

Database management systems are powerful tools for processing large volumes of structured, or normalized, data. Much of the data to be stored in computer systems, however, differs from normalized data in both its logical uses and the storage structure required for its effective management. For instance, Van Rijsbergen (1979) distinguishes database retrieval from information retrieval (IR)—the retrieval of references to text—by c ...

Results 1 - 6 of 6

Terms used **subword indexing**

Found **4** of **661,433**

Sort results by relevance

Display results expanded form

Save results to a Binder

Search Tips

Open results in a new window

Try an Advanced Search

Try this search in The Digital Library

Results 1 - 4 of 4

Relevance scale

1 Comparison of Word and Subword Indexing Techniques for Mandarin Chinese Spoken Document Retrieval

Hsin-Min Wang, Berlin Chen

October 2001 **Proceedings of the Second IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing**

Additional Information: full citation

2 Information Retrieval and Text Mining: Advances in phonetic word spotting

Arnon Amir, Alon Efrat, Savitha Srinivasan

October 2001 **Proceedings of the tenth international conference on Information and knowledge management**

Full text available: pdf(561.11 KB) Additional Information: full citation, abstract, references

Phonetic speech retrieval is used to augment word based retrieval in spoken document retrieval systems, for in and out of vocabulary words. In this paper, we present a new indexing and ranking scheme using metaphones and a Bayesian phonetic edit distance. We conduct an extensive set of experiments using a hundred hours of HUB4 data with ground truth transcript and twenty-four thousands query words. We show improvement of up to 15% in precision compare to results obtained speech recognition alone ...

3 Multimedia Information Processing: Automatic discovery of salient segments in imperfect speech transcripts

Dulce Ponceleon, Savitha Srinivasan

October 2001 **Proceedings of the tenth international conference on Information and knowledge management**

Full text available: pdf(117.97 KB) Additional Information: full citation, abstract, references

This paper addresses the problem of automatic detection of salient video segments for real-world applications such as corporate training based on associated speech transcriptions. We present a novel segmentation algorithm based on automatic speech recognition (ASR) applied to the audio track of the video. Our feature set consists of word n-grams extracted from the imperfect speech transcriptions. We use a two-pass algorithm that combines a boundary-based method with a content-based method. In th ...

4 Phonetic confusion matrix based spoken document retrieval

Savitha Srinivasan, Dragutin Petkovic

July 2000 **Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval**

Full text available: pdf(714.16 KB) Additional Information: full citation, abstract, references, citings, index terms

Combined word-based index and phonetic indexes have been used to improve the performance of spoken document retrieval systems primarily by addressing the out-of-vocabulary retrieval problem. However, a known problem with phonetic recognition is its limited accuracy in comparison with word level recognition. We propose a novel method for phonetic retrieval in the CueVideo system based on the probabilistic formulation of term weighting using phone confusion data in a Bayesian framework. We eval ...

Results 1 - 4 of 4

Your search matched **2** of **948708** documents.

A maximum of **2** results are displayed, **25** to a page, sorted by **Relevance** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again**.

phoneme classification and sequence [Search Again]

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

1 **The stochastic segment model for continuous speech recognition**
Ostendorf, M.; Digalakis, V.;
Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on , 4-6 Nov. 1991
Page(s): 964 -968 vol.2

[Abstract] [PDF Full-Text (408 KB)] **IEEE CNF**

2 **Automatic training of phoneme dictionary based on mutual information criterion**
Okawa, S.; Kobayashi, T.; Shirai, K.;
Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on , Volume: i , 19-22 April 1994
Page(s): I/241 -I/244 vol.1

[Abstract] [PDF Full-Text (224 KB)] **IEEE CNF**

Your search matched **1** of **950522** documents.

A maximum of **1** results are displayed, **25** to a page, sorted by **Relevance** in **descending** order.
You may refine your search by editing the current search expression or entering a new one the text box.
Then click **Search Again**.

phoneme classification and confusable

Search Again

Results:
Journal or Magazine = **JNL** Conference = **CNF** Standard = **STD**

1 **A neural tree network for phoneme classification with experiments on the TIMIT database**
Rahim, M.G.;
Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on , Volume: 2 , 23-26 March 1992
Page(s): 345 -348 vol.2

[Abstract] [PDF Full-Text (284 KB)] **IEEE CNF**

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 168 199 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
02.01.2002 Bulletin 2002/01

(51) Int Cl.7: G06F 17/30, G10L 15/00

(21) Application number: 01305315.2

(22) Date of filing: 19.06.2001

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 21.06.2000 GB 0015233

(71) Applicant: CANON KABUSHIKI KAISHA
Tokyo (JP)

(72) Inventors:
- Charlesworth, Jason Peter Andrew,
Canon Research
Bracknell, Berkshire RG12 2XH (GB)
- Garner, Philip Neil, c/o Canon Research
Bracknell, Berkshire RG12 2XH (GB)

(74) Representative:
Beresford, Keith Denis Lewis et al
Beresford & Co., 2-5 Warwick Court, High Holborn
London WC1R 5DH (GB)

(54) **Indexing method and apparatus**

(57) An indexing apparatus and method are described for use in identifying portions of data in a database for comparison with a query. In an embodiment, the index includes a key which comprises a sequence of phoneme classifications derived from the input query by classifying each of the phonemes in the input query with a number of phoneme classes, with the phonemes in each class being defined as those that are confusable with the other phonemes in the same class.

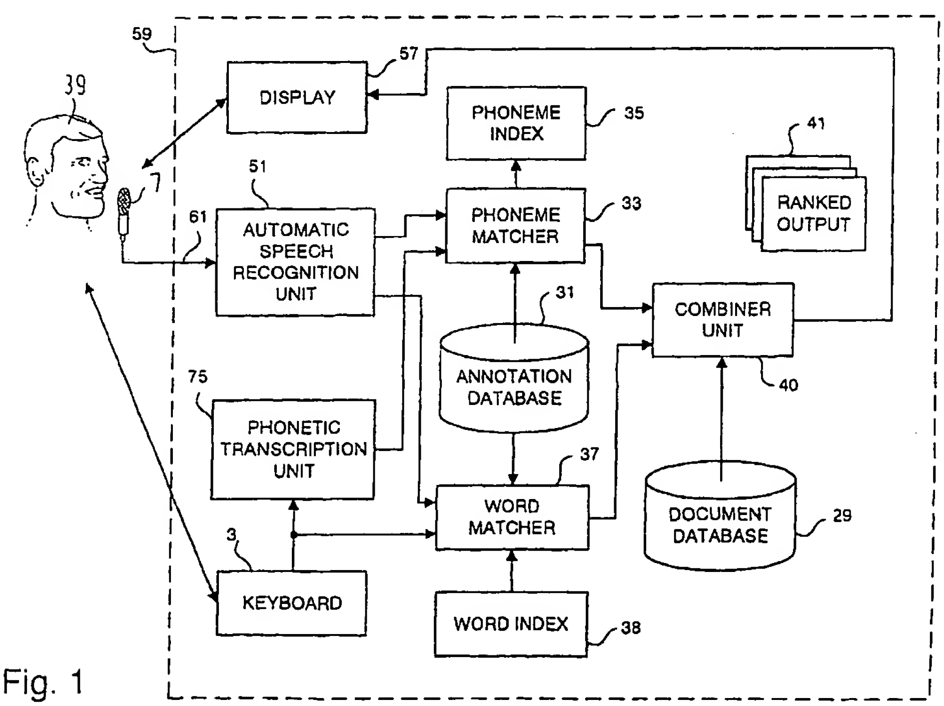


Fig. 1

EP 1 168 199 A2

US 20020120448A1

(19) United States
(12) Patent Application Publication (10) Pub. No.: US 2002/0120448 A1
Garner et al. (43) Pub. Date: Aug. 29, 2002

(54) SPEECH PROCESSING SYSTEM

(76) Inventors: Philip Neil Garner, Tokyo (JP); Jason Peter Andrew Charlesworth, Haslemer (GB)

Correspondence Address:
FITZPATRICK CELLA HARPER & SCINTO
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112 (US)

(21) Appl. No.: 09/986,914

(22) Filed: Nov. 13, 2001

(30) Foreign Application Priority Data

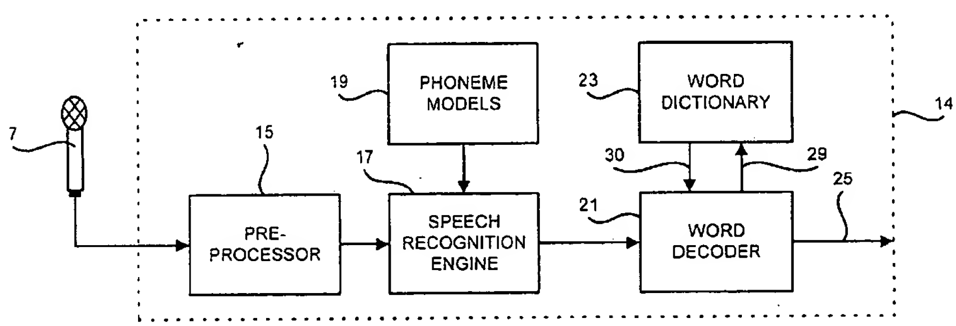
Nov. 20, 2000 (GB) 0028277.2

Publication Classification

(51) Int. Cl.^{7} G10L 15/04
(52) U.S. Cl. 704/254

(57) ABSTRACT

A system is provided for decoding one or more sequences of sub-word units output by a speech recognition system into one or more representative words. The system uses a dynamic programming technique to align the sequence of sub-word units output by the recognition system with a number of dictionary sub-word unit sequences representative of dictionary words to identify the most likely word or words corresponding to the spoken input.



US 20020120447A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2002/0120447 A1

Charlesworth et al. (43) Pub. Date: Aug. 29, 2002

(54) SPEECH PROCESSING SYSTEM

(76) Inventors: Jason Peter Andrew Charlesworth, Surrey (GB); Jebu Jacob Rajan, Bracknell (GB)

Correspondence Address:
FITZPATRICK CELLA HARPER & SCINTO
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112 (US)

(21) Appl. No.: 09/985,543

(22) Filed: Nov. 5, 2001

(30) Foreign Application Priority Data

Nov. 7, 2000 (GB) 0027178.3

Publication Classification

(51) Int. Cl.⁷ G10L 15/04

(52) U.S. Cl. 704/254

(57) ABSTRACT

A system is provided for allowing a user to add word models to a speech recognition system. In particular, the system allows a user to input a number of renditions of the new word and which generates from these a sequence of phonemes representative of the new word. This representative sequence of phonemes is stored in a word to phoneme dictionary together with the typed version of the word for subsequent use by the speech recognition system.